Synar Survey Estimation System

Version 2.1

User Manual

July 30, 2004

Prepared by the

Table of Contents

SYNAR SURVEY ESTIMATION SYSTEM MANUAL

1.   Introduction

The Synar Survey Estimation System (SSES) is an optional software tool developed by the Center for Substance Abuse Prevention (CSAP) to assist states in estimating and reporting of their annual Synar survey results. The system is a Visual Basic application designed to operate in Microsoft Excel 97 or higher version. SSES is menu-driven software; however, in order to use it, the state must create a micro-level data file, containing a record for each sampled tobacco outlet. This data file will be in the form of an Excel spreadsheet and must be formatted in exact accordance with the specifications provided below in order for the application to run correctly. Once the data file is set up, running the estimation system is as simple as selecting the appropriate option from the drop-down menu.

New features were added in this Version 2.1. The SSES now produces all Synar survey results required in the Annual Synar Report (ASR) including all mandatory Forms. Because of this significantly increased capability, the system requires some additional data beyond what Version 1.1 required, notably information on youth inspectors. However, it will reduce reporting burden substantially by enabling the user to simply attach the SSES output tables to the ASR. In addition, a new sample size calculator has been added which calculates the appropriate sample size for the State's Synar survey. It is designed to be used prior to conducting the survey.

The following guidelines and explanations are designed to help the user understand both how to utilize the software and how to interpret the output it produces.

2.   Setting up the Software

The SSES2.1 software program is available in CD-ROM, which also contains other peripheral files. Follow the step-by-step instructions below to install the software in your computer.

a.   Create a SSES directory in your computer to store the program.
b.   Insert the SSES2.1 CD-ROM in the CD drive.
c.   Copy all the files in the CD-ROM to the SSES directory or at least the following two files: the software program file named "SSES_V21.xla" and the PDF file "SSESManualV21.pdf" that contains the manual. (Then the entire manual will be available in the Help menu under the SSES system menu.)

This completes the installation. The SSES2.1 CD-ROM also includes data entry templates, which are useful in preparing a data file needed to run the system. There are also some example data sets that can be used to try out the system. The templates and example data files can be stored separately in another directory.

3.   Invoking the System

Since the system is developed as a Microsoft Excel add-on, the system is invoked by simply opening the saved program file, SSES_V21.xla in Excel. This can be achieved in two ways as follows:

a.   Open Excel in your computer and then using the file menu of Excel, open the program file SSES_V21.xla.

b.   Or open the Windows Explorer and locate the program file SSES_V21.xla. Double click over the file, and then it will be opened in Excel.

At the moment the system is invoked, a pop-up window will appear as shown below.
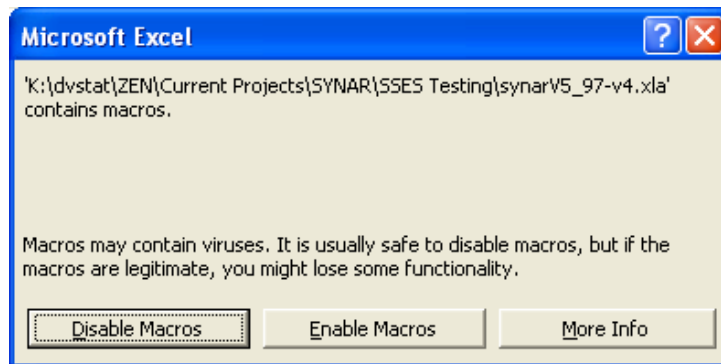
Select the "Enable Macros" button in the pop-up window. If your computer has Windows XP and you cannot open it by selecting the "Enable Macros" button because of a higher security setting, then you have to lower the security level of your Excel by following the steps laid out below.

a.   From the **Tools** menu, select **Options**.

b.   Click the **Security** tab.

c.   Press the **Macro Security** button (lower right), which brings up another window.

d.   Select the **Medium** security level and press **OK**.

e.   Press **OK** again to close the window.

f.	Open the SSES application and then select the "Enable Macros" button. When enabling macros is successfully done, a pop-up window, "About SSES" will appear as shown below.
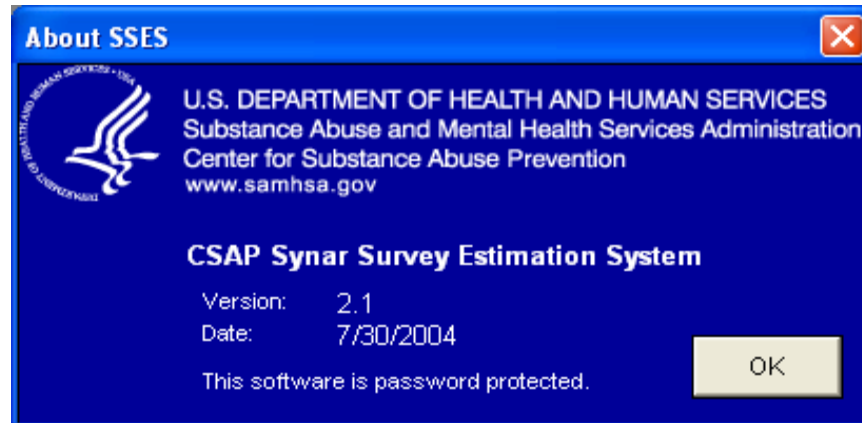
Click **OK**. You will now see the "CSAP SYNAR" menu bar on your Excel menu panel. Click the menu, you will see the following options:

c.	**Stratified SRS**: a Synar survey data analysis option, which is suitable to analyze the Synar survey data collected using a stratified simple (or systematic) random sample (SRS) design. It has two sub-options, "With FPC" and "Without FPC."

d.	**Stratified CLUSTER**: a Synar survey data analysis option, which is suitable to analyze the Synar survey data collected using a stratified cluster sample design. It has two sub-options, "With FPC" and "Without FPC."

e.	**Sample Size Calculator**: this is an independent module, which is expected to be used at the beginning of the survey. It calculates the sample sizes for the Synar survey before selecting a sample.

f.	**Help**: this feature is used to browse specific topics or the entire user manual.

FPC stands for finite population correction. The With-FPC option applies such correction in the variance estimation and the other option does not. Section 4.2 provides advice on how to choose an option depending on the desired analysis.
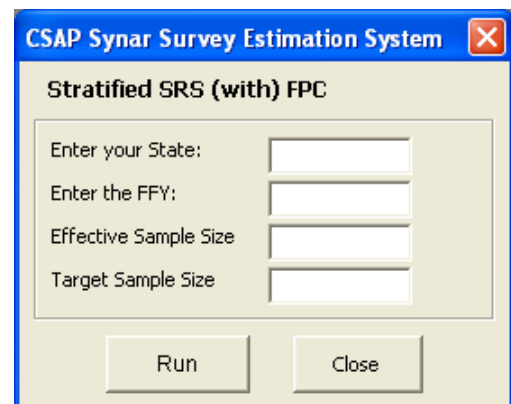
4.   Running the System

To run a menu option, it is necessary to provide data to the system. Data requirements are vastly different for the different options, and so separate sections of this manual are devoted to discussion for each of the options: Section 5 for the Stratified SRS option, Section 6 for the Stratified CLUSTER option. The system options will only run properly if the data file is in the correct format, as specified in these sections, and, therefore, it is essential that you read the relevant sections carefully. Section 4.1 provides a brief overview of running the system with a prepared dataset. You can make a test run of the system using a pre-prepared example data set that came with the CD-ROM or with your own dataset prepared according to the instructions in Sections 5 and 6.

4.1   Running an Analysis Option

To run an analysis option, after invoking SSES as explained in Section 3, follow the steps given below:

    a.   Using the Excel file menu, open an Excel Synar data file, prepared according to the specified format for the analysis (see Sections 5 and 6 for the specified file formats).

    b.   Select an analysis menu option (either the Stratified SRS or CLUSTER) appropriate for the data file (the example data set names contain "SRS" or "CLUSTER" signifying the sample design).

    c.   For each of the analysis menu option, you have two choices: "With FPC" and "Without FPC." Choose one of them (see Section 4.2 for more information).

    d.   When you click your choice, a pop-up window as shown on the right appears. At the prompt enter:

- State name or abbreviation.
- Federal Fiscal Year (FFY) – this is usually one year ahead of the calendar year of the survey (e.g., if the survey is conducted in 2004, FFY is 2005).
- Effective sample size - this is the sample size needed to meet the SAMHSA precision requirement assuming that simple random sampling (SRS) design is used.
- Target sample size – this is the sample size needed to achieve the desired precision requirement with a complex sample design if the actual sample design is more

complex than SRS. To obtain the target (or sometimes called minimum) sample size, multiply the effective sample size by the design effect (for more discussion for the sample size determination procedure, see the most recent version of Synar Regulation: Sample Design Guidance).

e. Click **Run**, and then the system prompts you to check the entered sample sizes as shown below.
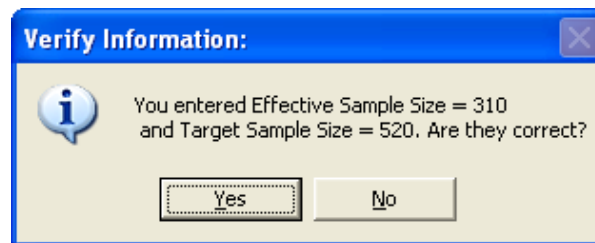
f. If correct, click **Yes**; otherwise, click **No** and correct the values. Click **Run** again.

g. If there is no unacceptable error, the system reads the data file and provides you with the number of records read (see below).
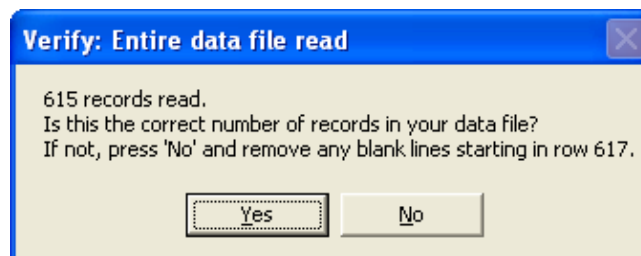
h. If the number is correct, click **Yes**; otherwise, click **No** and check the data set to determine if any blank row is inserted by mistake.

i. If you click yes, the system runs and produces four SSES output tables (see Section 7 for discussion of the output tables).

j. In step g, if the data file contains some unacceptable errors, the system will generate appropriate error message (see Appendix A for error messages and troubleshooting). You have to correct the errors before you are able to run the system.

4.2   Choosing an Analysis Option

The main criterion for choosing an analysis option is the survey sample design. If outlets are selected directly by census or simple or systematic random sampling with or without stratification, you should choose the Stratified SRS option. If clustering was used either with or without stratification, select the Stratified CLUSTER option. (Stratification means explicit stratification unless specified otherwise).

If no stratification is used, the whole state is treated as a single stratum, and, therefore, the qualifier "Stratified" applies to both sample designs with and without stratification.

For each of these main analysis options, there are two sub-options: "With FPC" or "Without FPC." The abbreviation "FPC" stands for finite population correction, which is applicable only for variance estimation purposes when a sample is selected from a finite population by a without-replacement sampling method. All State Synar surveys fall under this category, and, therefore, the FPC is technically applicable to all of them. This sampling method is assumed in this manual unless otherwise specified.

However, sometimes FPC should not be used. The choice largely depends on the particular sample design. More discussion on this will be given in Sections 4.2.1 and 4.2.2 when discussing each of the two main analysis options.

4.2.1　Stratified SRS (for Designs without Clustering)

This option is applicable for the following sample designs:

a. Census;
b. Simple random sample of outlets with or without stratification;
c. Systematic random sample of outlets with or without stratification.

If outlets are selected state-wide by census or SRS without stratification, then the whole state becomes the single stratum, which should be noted as such in the data file.

The FPC is always applicable and appropriate. Applying it with the "With FPC" option in SSES reduces the variance estimate, generally making it beneficial for the states to use the With-FPC option. However, when the FPC is negligible, the difference between the variance estimates produced by the two approaches is small - this happens when the sampling rate is small (the rule of thumb is 10 percent). In this case, the Without-FPC option may be used even for without-replacement sampling to obtain a slightly conservative variance estimate – a variance estimate is said to be conservative when it is positively biased, that is, the variance estimate is slightly larger than the one obtained with the With-FPC option. Slight conservatism is often opted for in practice so as not to overstate the precision of the estimated RVR. However, the choice is up to the user in this case, and either choice is acceptable. (For discussion on census surveys, see the technical remark below in this section.)

The variance formula used in the system for the Stratified SRS option is for simple random sampling, and there is no separate option for systematic sampling. There is not an explicit variance formula for systematic sampling, although an approximate estimate can be produced by other means. If systematic sampling was done with a randomized list, then systematic sampling is equivalent to simple random sampling, and the option is valid. However, if the list is sorted before sampling to achieve implicit stratification, the Stratified SRS option provided by the system does not reflect the implicit stratification in variance estimation, and, therefore, it is likely to produce a conservative variance estimate since the implicit stratification is ignored. If you want to reflect the implicit stratification in variance estimation, it can be done to a certain extent by using the post-stratification approach, which is discussed in Section 9.6. In either case, it is advisable to use the With-FPC option to reduce the conservatism of the variance estimate.

Some states use post-stratification with a state-wide SRS to produce more efficient RVR estimates or separate RVR estimates for sub-state areas. This Stratified SRS option is suitable to implement such estimation strategy by defining the post-strata as the variance strata (for more detailed discussion, see Section 9.6).

The table below provides a summary of the discussion in this section.

Table 1.   Summary table for the Stratified SRS option

Sample Design	Stratum	Sampling Rate	FPC Option	Comment
State-wide SRS/Census	State	10% or higher	With	
State-wide SRS	State	Lower than 10%	Both	"Without" optional
Stratified SRS	Design strata	10% or higher	With	
Stratified SRS	Design strata	Lower than 10%	Both	"Without" optional
State-wide Systematic[1]	State	Any	With	Post-stratification useful
Stratified Systematic[1]	Design strata	Any	With	Post-stratification useful

[1] Note: systematic sampling from a sorted list.

Technical remark: Even when a census survey is conducted, it is rare to achieve 100 percent completion. If the completion rate is reasonable, the RVR estimate would be very precise but it is not completely error free. If it can be assumed that noncompletion occurs randomly, the resulting sample from census can be treated as an SRS with a high sampling rate. For example, if the completion rate is 90 percent and noncomplete outlets occur haphazardly, then the resulting final sample from the census can be treated as an SRS with 90 percent sampling rate. Then, such sample data can be analyzed using the Stratified SRS with the With-FPC option (it is very important to use the With-FPC option in this case).

4.2.2   Stratified CLUSTER (for Designs with Clustering)

This option is applicable for any cluster sample designs with single or multiple stages with or without stratification. If no explicit stratification is used, then the whole state becomes the single stratum, which should be noted as such in the data file.

Unlike the Stratified SRS case, selecting one of With-FPC and Without-FPC options must be done with care. Although the sample design is always without-replacement sampling of primary sampling units (PSUs), the With-FPC option for a cluster sample design could lead to a biased variance estimate depending on a number of factors. Stratification is not an important factor, and, thus, it is not included in the following discussion, where each case is dealt with separately:

a. **One- or two-stage cluster design with PSUs sampled by equal probability sampling (e.g., SRS) without replacement**: the With-FPC option should be used and an unbiased variance estimate will be obtained.

b. **One- or two-stage cluster design with PSUs sampled by probability proportional to size (PPS) or other unequal probability sampling without replacement**: different treatments are required for different situations.

   - *No certainty PSU (i.e., all PSU sampling probabilities are less than one)* - the With-FPC option is recommended in most cases. The variance estimate can be somewhat liberal but it is acceptable if not excessive. If the overall sampling rate is small (< 10 percent), it is recommended to use the Without-FPC option, which will produce somewhat conservative variance estimate, which is more preferable than a liberal one.

   - *Some certainty PSUs (i.e., some PSUs are selected with a probability of one)* - the certainty PSUs should be treated differently from other non-certainty PSUs by forming a separate variance stratum or strata for them. For one- or two-stage designs, all certainty PSUs should be put in a single variance stratum and then the PSUs should be used as the variance units (see Section 9.4 for details). Then use the With-FPC option.

c. **Three- or higher stage cluster design**: always use the Without-FPC option. If there are some certainty PSUs, each of them should be made a separate variance stratum of its own and then the secondary sampling units in the certainty PSUs should form the variance units (see Sections 9.4 and 9.7 for details).

The above discussion is summarized in the following table:

Table 2. Summary table for the Stratified CLUSTER option

Cluster Sample Design	PSU sampling	Sampling Rate	FPC Option	Comment
1- or 2-stage	Equal	Any	With	Unbiased variance
1- or 2-stage without certainty PSU	Unequal	Lower than 10%	Without	Conservative / "With" optional
1- or 2-stage without certainty PSU	Unequal	10% or higher	With	Slightly liberal / Without – too conservative
1- or 2-stage with certainty PSU	Unequal	Any	With	Put certainty PSUs in a single variance stratum
3- or higher-stage with or without certainty PSU	Unequal	Any	Without	Make each certainty PSUs a variance stratum

Technical remark 1: The system treats even a one-stage design as a two-stage design with 100 percent second stage sampling. This is technically correct. However, because of the occurrence of eligible noncompletion, the one-stage design often becomes actually a two-stage design.
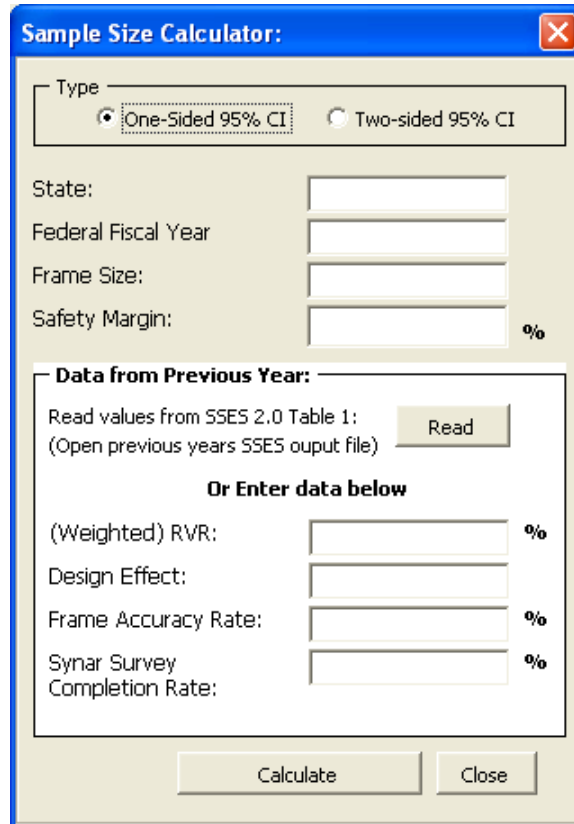
Technical remark 2: When PSUs are selected with unequal probabilities without replacement, besides the probabilities of selecting individual PSUs, the probability that any two sample PSUs are selected jointly within a stratum (called the joint probability) is needed to obtain an unbiased variance estimate. However, the joint probabilities are difficult to calculate and not usually available, and that is the working assumption of the system. Therefore, an approximate approach to variance estimation must be used as discussed above. Even though it is rare in practice, if sampling of PSUs has been done with replacement, the Without-FPC option provides an unbiased variance estimate. This option should also be used for a without-replacement design with more than two stages of sampling, and with a small overall sampling rate. This option produces an unbiased variance estimate for with-replacement sampling of PSUs but a conservative variance estimate for without-replacement sampling of PSUs. It is generally preferable to be conservative rather than liberal in order not to overstate the precision of the RVR estimate. Therefore, if you are in doubt, use the Without-FPC option. You can try both, and if you get similar variance estimates, you can use the Without-FPC option without being too conservative. Stratifying the PSUs by their size is helpful in reducing the bias in the variance estimate (with the With-FPC option) because it makes the within stratum PSU probabilities more equal.

4.3   Running the Sample Size Calculator

This is a new module added to Version 2.1 of the system. It is an independent module to be used at a different time point than the others, i.e., just before the Synar sample is selected. To run the Calculator, follow the steps laid out below.

      a.   Invoke the SSES by opening the program file in Excel or double clicking the file name in Windows Explorer (see Section 3 for a more detailed invoking procedure).

b. Select the Calculator option from the CSAP SYNAR menu options, and then a pop-up
    window will appear as shown below.

c. Select the type of confidence interval for which the SAMHSA precision requirement of
    3 percent margin of error should be applied. Choices are:
    *   *One-sided 95% CI* – this is the default choice and the same precision requirement
        can be met with a smaller sample size than with the two-sided choice.
    *   *Two-sided 95% CI*

d. Provide four items needed in the upper part of the pop-up window, including :
    *   *State name* (full or abbreviation);
    *   *Federal Fiscal Year (FFY)* – this is usually one year ahead of the calendar survey
        year;
    *   *Frame Size* – the size of the list sampling frame from which a Synar sample is to be
        drawn.

- *Safety Margin* – a percentage figure between 0 and 50 by which the sample size should be inflated to make sure the sample size is big enough. If 0 is given, no inflation is made.

e. Provide information needed in the lower part of the pop-up window. This can be done in two ways: (1) if SSES2.1 was used to analyze the previous Synar survey data the needed information can be read in from the SSES output file. To do that, first close the pop-up window and open the SSES output Table 1 of the last year's survey, and then select the Sample Size Calculator again. In the pop-up window, click **Read**. This method cannot be used if SSES2.1 was not used previously. (2) type in the required information:

- *(Weighted) RVR* – last year's RVR in percentage;
- *Design Effect* – design effect estimated from last year's survey or by other appropriate means (when a new design is used);
- *Frame Accuracy Rate* – frame accuracy rate in percentage estimated from last year's survey or other appropriate estimate (if a new frame is used);
- *Synar Survey Completion Rate* – the completion rate in percentage experienced last year.

f. Click **Calculate**, and then the output is immediately displayed in a new worksheet.

g. Move away or Close the pop-up window (the entered information is still kept) to see the output table (see Exhibit 1 for an example).

h. If satisfied, save the output; otherwise reopen the pop-up window by selecting the Sample Size Calculator menu option for another run with a modified input.

The output table provides three sample sizes: effective, target, and planned original. The planned original sample size is the one to be used to select a sample. The actual sample size can be somewhat different from the planned one. For example, if a cluster design is used with varying sizes of PSUs, it is often not possible to exactly control the sample size. However, every effort should be exerted to select a sample close to the planned original sample size. It is required to report various sample sizes in the ASR, three of which are the effective, target, and realized (opposed to planned) original sample sizes. When you run an analysis option to analyze the Synar survey results, you must provide the effective and target sample sizes at the beginning to the run where you can report the sample sizes obtained by the Calculator here.

Technical remark: If a new sample design, which is very different from the one used last year, is going to be used, the last year's design effect is not applicable. A rough estimate of the design effect for the new design can be made to use it in the Calculator instead of the last year's value. See the recent version of the Synar Regulation: Sample Design Guidance booklet published by SAMHSA for more information about estimating the design effect for a new sample design.

Exhibit 1. Sample Size Calculator output table – an example

5. **Data Requirements and Preparation for the Stratified SRS Option**

It is important that someone who has good knowledge of the actual survey design either creates or oversees the creation of the input data file for the survey data analysis options. The input data file must be an Excel file in a form of worksheet.

Create an Excel worksheet containing the 12 designated column titles from Columns A to L in the first row without any blank columns in between, in the exact order as listed below.

    A. Outlet ID

    B. Sampling Stratum

    C. Total Outlets in Sampling Stratum

D. Variance Stratum

E. Total Outlets in Variance Stratum

F. Response Disposition Code

G. Violation Flag

H. Outlet Type

I. Youth Inspector ID

J. Youth Inspector Gender

K. Youth Inspector Age

L. VM Frame Size in Sampling Stratum

The first row of the Excel data worksheet must be used for the column title. The actual survey data must start in Row 2 of the worksheet and continue in consecutive rows. The easiest way to start is to use the data entry template ("SSES Data Entry Template-SRS.xls") provided with the system. Keep the following points in mind while preparing the data file.

a. Any alphanumeric names can be used as the column headings; however, it is helpful to use easily understood column titles. Blank or numeric names will not be accepted.

b. There should be one row for each sample outlet. All sample outlets should be included whether they are eligible or ineligible, inspection completed or noncompleted. If a reserve or replacement sample was used, all outlets from both the original sample and the reserve or replacement sample that was used (fielded) should be in the data file.

c. No blank row in the data set is allowed, and, therefore, the number of data lines excluding the title row must be exactly the same as the original sample size.

d. No distinction is recognized between upper and lower cases in the alphabetical data.

e. There must not be any hidden or blank columns in between Columns A to L.

f. Column M must be blank for the Stratified SRS options; otherwise the system will assume that the data file is for the Stratified CLUSTER options. However, the other columns are available for any other uses.

g. Save the file frequently while entering the data.

The system recognizes a blank row in the data file as the end of the data. After reading the data file, the system provides a message giving the number of lines read. If a blank row has been

imbedded by mistake, the number of data lines read will be different from the expected number, and such a mistake must be corrected before running the system.

Details on the data requirements for each of the 12 columns follow.

Column A - Outlet ID

A unique identifier for each outlet. This field can be numeric or character but should be unique and non-missing. The system checks whether the ID is unique or not and missing or not, and sends an error message otherwise.

Column B - Sampling Stratum

The sampling stratum to which the outlet belongs. There must be an unique identifier for each sampling stratum. All outlets in the same stratum must have the same identifier in Column B. This data field can be numeric or character but must not be missing. If vending machines were selected separately from over-the-counter outlets, then all vending machines should be placed in separate strata of their own. If no stratification was used, simply enter the same identifier in every row. In this case, the entire state is regarded as a single stratum.

Column C - Total Outlets in Sampling Stratum

The total number of outlets in the sampling frame that are in the sampling stratum. This field must be numeric and its value must be identical for all sample outlets within the same sampling stratum. If no stratification was used, the whole state becomes the single stratum, and so enter the total number of outlets in the sampling frame in every row. The information provided in this column is used to produce frame counts in SSES output Table 2 (see Exhibit 3 in Section 7.2).

Column D - Variance Stratum

An identifier for each variance stratum. This field can be numeric or character but must not be missing. In most cases the variance stratum will be the same as the sampling stratum, and it is advantageous to use the same ID as the sampling stratum in this case— just copy the ID's in Column B over to Column D. However there are some situations in which they may differ. If there is only one eligible responding or ineligible outlet in a sampling stratum (and it is not the only outlet in the sampling frame in that stratum), then this sampling stratum must be combined with one or more other sampling strata to form a valid variance stratum with more than one such outlet. See Section 9.2 for guidelines. If no stratification was used, simply enter the same identifier in every row. If vending machines were selected separately from over-the-counter outlets, then all vending machines should be placed in separate strata of their own.

Column E – Total Outlets in Variance Stratum

The total number of outlets in the sampling frame that are in the variance stratum. This field must be numeric and its value must be identical for all outlets within the same variance stratum. The information provided in this column is used to calculate the sampling weights and, thus, it affects the whole estimation process the system performs. Therefore, it is extremely important to provide the correct values in this column. If the sampling strata are the same as the variance strata, then this column should be identical with Column C, where the same information for the sampling strata is provided. In this case, the system checks the equality of the values in Columns C and E if the same IDs are used for the sampling strata and variance strata. However, if different IDs are used, the system treats as if variance strata were redefined and does not check the equality of the values given in the two columns (C and E). If two or more sampling strata have been combined to form a variance stratum, the frame size for that variance stratum should equal the sum of the frame sizes in the combined sampling strata. On the other hand, if a sampling stratum is split into a number of variance strata, the frame size for that sampling stratum should equal the sum of the frame sizes in the variance strata defined within that sampling stratum. The system does not check this consistency, and so you should make sure that the data file has internal consistencies in regard to the variance strata.

Column F - Response Disposition Code

A code to indicate the final response disposition of the outlet inspection. This field must be an alphanumeric code exactly as specified in Table 3. If, for any records, this field is coded as "other" (i.e., N9 or I10), then you should give other reasons and counts in the space provided in SSES output Table 3 (Synar Survey Sample Tally Summary) – more discussion is given in Section 7.3.

Table 3.   Response disposition codes

Status	Disposition code	Definition
Eligible complete	EC	Eligible and inspection complete outlet
Eligible noncomplete	N1	In operation but closed at time of visit
Eligible noncomplete	N2	Unsafe to access
Eligible noncomplete	N3	Presence of police
Eligible noncomplete	N4	Youth inspector knows salesperson
Eligible noncomplete	N5	Moved to a new location but not inspected.
Eligible noncomplete	N6	Drive thru only/youth inspector has no driver license
Eligible noncomplete	N7	Tobacco out of stock
Eligible noncomplete	N8	Run out of time
Eligible noncomplete	N9	Other noncompletion reason(s)

Table 3.   Response disposition codes (continued)

Status	Disposition code	Definition
Ineligible	I1	Out of business
Ineligible	I2	Does not sell tobacco products
Ineligible	I3	Inaccessible by youth
Ineligible	I4	Private club or private residence
Ineligible	I5	Temporary closure
Ineligible	I6	Unlocatable
Ineligible	I7	Whole sale only/Carton sale only
Ineligible	I8	Vending machine broken
Ineligible	I9	Duplicate
Ineligible	I10	Other ineligibility reason(s)

Column G - Violation Flag

An indicator of whether or not the outlet was found to be in violation. This field must be 0, 1, or blank. Table 4 specifies the coding. Outlets found to be in violation must have this flag set to "1". Outlets found to be in compliance must have this field set to "0" or blank**.** Ineligible or eligible noncomplete outlets cannot have this field set to "1". Use the following violation codes for inspection results:

Table 4.   Violation codes

Inspection result	Violation code
Outlet found to be in violation	1
Outlet found to be in compliance	0 or blank
Eligible noncomplete outlet	blank
Ineligible outlet	blank

Column H – Outlet Type

An indicator of the type of the outlet. There are only two types of outlets: Over-the-counter coded as "OTC" and vending machine coded as "VM". If the type is unknown, it must be coded as "UNK"; a missing value is not allowed. When a record with unknown outlet type is present in the data file, a special message will be printed at the bottom of SSES output Table 2 indicating that the overall counts may not be equal to the sum of OTC and VM counts.

Column I – Youth Inspector ID

A unique identifier for each youth inspector. This data field can be numeric or character. Each inspector must have a unique ID, and for a given inspector ID, inspector characteristics such as

gender and age must be consistent. The system does not utilize this field if the outlet record has a response disposition code of ineligible or eligible noncomplete.

Column J – Youth Inspector Gender

This data field should be coded with the youth inspector's gender, either "M" for male, "F" for female, or blank if unknown. Other values are not allowed. If the gender is unknown but the age is within the allowable range, then the inspector will be categorized as "Other." In this case, a space will be provided at the bottom of SSES output table 4, where you should give explanation for the "Other" category.

Column K – Youth Inspector Age

Youth inspector age should be numeric or blank if unknown. If a sample outlet is eligible and has a complete inspection (i.e., the record disposition code is "EC"), then the age must be greater than or equal to 14 but less than or equal to 18. If the age is outside of this range or unknown, the outlet should be declared as eligible noncomplete because the inspection is invalid and then assigned the disposition code "N9", for "Other" noncompletion. The user should give this specific reason in a space provided at the bottom of SSES output Table 4.

For a given Youth Inspector ID, the gender and age must be consistent across all records; otherwise, an error message will pop up.

Column L – VM Outlet Frame Size in Sampling Stratum

This field provides the total number of vending machine (VM) outlets in the sampling frame that are in the stratum. This field must be numeric or blank and its value must be identical for all outlets within the same sampling stratum. If it is unknown or no VM is selected, leave it blank.

Sampling weights are not needed because the system computes them from the information provided. It also performs nonresponse adjustment at the stratum level.

6.   Data Requirements and Preparation for the Stratified CLUSTER Option

It is important that someone who has good knowledge of the actual survey design either creates or oversees the creation of the input data file for the survey data analysis options. The input data file to SSES must be an Excel file in a form of worksheet.

Create an Excel file containing the 18 designated column titles from Columns A to R in the first row without any blank columns in between, in the exact order as listed below.

A. Outlet ID

B. Sampling Stratum

C. Total PSUs in Sampling Stratum

D. Sampling PSU

E. Total Outlets in Sampling PSU

F. Variance Stratum

G. Total Variance Units in Variance Stratum

H. Variance Unit

I. Total Outlets in Variance Unit

J. Sampling Weight

K. Response Disposition Code

L. Violation Flag

M. Total Outlets in Sampling Stratum

N. Outlet Type

O. Youth Inspector ID

P. Youth Inspector Gender

Q. Youth Inspector Age

R. VM Outlet Frame Size in Sampling Stratum

You must use the first row of the Excel data worksheet for the column title. The actual survey data must start in Row 2 of your worksheet and continue in consecutive rows. The easiest way to start is to use the data entry template ("SSES Data Entry Template-CLUSTER.xls") provided with the system. Keep the following points in mind while preparing the data file.

a. Any alphanumeric names can be used as the column headings; however, it is helpful to use easily understood column titles. Blanks or numeric names are not accepted.

b. There should be one row for each sample outlet. All sample outlets should be included whether they are eligible or ineligible, inspection completed or noncompleted. If a reserve or replacement sample was used, all outlets from both the original sample and the reserve or replacement sample that was used (fielded) should be in the data file.

c. No blank row in the data set is allowed, and, therefore, the number of data line excluding the title row must be exactly the same as the original sample size.

d. No distinction is recognized between upper and lower cases in the alphabetical data.

e. There must not be any hidden or blank columns in between Columns A to R. The other columns are available for ant other uses.

f. Save the file frequently while entering the data.

The system recognizes a blank row in the data file as the end of the data. While reading the data file by the system, when a blank row is hit, the system gives a message, which provides the number of data lines read in from just before the blank row and asks you if it is correct. If a blank row has been imbedded by mistake, you will notice that the number of data lines read is different from the expected number, so that you can investigate and correct such a mistake before running the system.

Details on the data requirements for each of the 18 columns follow. The following example will be used to explain some of the data fields to be included in the data file.

Example: In a sampling stratum named STR100, 12 PSU clusters are created and three of them selected in the sample. These selected PSUs are identified by 101, 102, and 103, and they have 34, 27, and 20 outlets in the sampling frame within the PSUs, respectively, before selecting outlets from them for the survey sample. From each of these PSUs, 10 outlets are selected, and thus, altogether 30 outlets are selected from the sampling stratum STR100. These outlets are uniquely identified by 10101, 10102, …,10110 for the 10 sample outlets from PSU 10, by 10201, 10202,…, 10210 for the 10 outlets selected from PSU 102, and by 10301, 10302, …, 10310 for the 10 outlets from PSU 103. This example is summarized in the following table.

Sampling stratum ID	Total PSUs in sampling stratum	Number of PSUs selected	Sample PSU ID	Number of outlets in sample PSU	Number of sample outlets in sample PSU
STR100	12	3	101	34	10
STR100	12	3	102	27	10
STR100	12	3	103	20	10

Details on the data requirements for each of the 18 fields follow.

Column A - Outlet ID

A unique identifier for each outlet. This field can be numeric or character but should be unique and nonmissing. The system checks the uniqueness and missingness of the ID. For the example above, outlet ID's for the 30 sample outlets from STR100 are defined as 10101, 10102, …, 10110, 10201, 10202, …, 10210, 10301, 10302, …, 10310.

Column B - Sampling Stratum

The sampling stratum to which the outlet belongs. There must be a unique identifier for each sampling stratum. All outlets in the same stratum will have the same identifier in Column B. This field can be numeric or character but must not be missing. If vending machines were selected separately from over-the-counter outlets, then all vending machines should be placed in separate strata of their own. If no stratification was used, simply enter the same identifier in every row. In the above example, the 30 data lines for the 30 sample outlets would be given "STR100" for this field.

Column C - Total PSUs in Sampling Stratum

The total number of PSUs in the sampling frame that are in the sampling stratum. This field must be numeric and its value must be identical for all outlets within the same sampling stratum. For the above example, this field should contain "12" for the 30 sample outlets. If no stratification was used, enter the total number of PSUS in the sampling frame in every row.

Column D - Sampling PSU

An identifier for each sampling PSU. This field can be numeric or character. For the example above, the 10 data lines for the 10 sample outlets from PSU 101 would have "101" for this field, the 10 sample outlets from PSU 102 would have "102," etc.

Column E - Total Outlets in Sampling PSU

The total number of outlets in the sampling frame that are in the sampling PSU. This field must be numeric and its value must be identical for all outlets within the same sampling PSU. For the example above, the 10 records from PSU 101 would have this field filled by "34", those from PSU 102 by "27", and those from PSU 103 by "20".

Column F - Variance Stratum

An identifier for each variance stratum. This field can be numeric or character but must not be missing. In most cases the variance stratum will be the same as the sampling stratum, and it is

advantageous to use the same ID as the sampling stratum in this case – just copy the ID's in Column B over to Column F. For the above example, the 30 sample outlets would have "STR100" for this field, if no redefinition of the variance strata and the same sampling stratum ID is used. However there are some situations in which the variance strata may differ from the sampling strata. For example, if there is only one PSU in a sampling stratum (and it is not the only PSU in the frame in that sampling stratum), then this sampling stratum must be combined with one or more other sampling strata to form a valid variance stratum with more than one PSU (i.e., variance unit). See Section 9.3 for guidelines. If no stratification was used, simply enter the same identifier in every row regarding the whole state as a single stratum. If vending machines were selected separately from over-the-counter outlets, all vending machines should be placed in separate strata of their own.

Column G - Total Variance Units in Variance Stratum

The total number of variance units in the sampling frame that are in the variance stratum. This field must be numeric and its value must be identical for all outlets within the same variance stratum. If the variance strata are defined as the same as the sampling strata and the same IDs are used, then the system checks whether Column G values are the same as Column C values. In this case, the 30 outlets in the above example will have "12" for this field. If two or more sampling strata have been combined to form a variance stratum, the total number of variance units for that variance stratum should equal the sum of the total number of PSUs in the combined sampling strata. On the other hand, if a sampling stratum is split into a number of variance strata, the sum of the numbers of variance units in the split variance strata should be equal to the total number of PSUs in the sampling stratum. The system does not check this kind of consistency, and so you should make sure that the data file has internal consistencies in regard to the variance units.

Column H - Variance Unit

An identifier for each variance unit. This field can be numeric or character. In most cases the variance unit will be the same as the sampling PSU. It is also easier to prepare the data file if the same ID is used. If so, for the above example, the 30 outlets will have Variance Unit ID equal to 101, 102, or 103 depending on their PSU ID. However there are some situations in which they may differ. If there is only one eligible responding or ineligible outlet in a sampling PSU (and it is not the only outlet in the frame in that PSU), then this PSU must be combined with one or more other PSUs to form a valid variance unit, unless the "Without-FPC" option is used. When the "Without-FPC" option is used, however, single outlet PSUs are allowed. See Section 9.3 for a further explanation of this issue. For the case of certainty PSUs see Section 9.4.

Column I - Total Outlets in Variance Unit

The total number of outlets that are in the variance unit in the frame. This field must be numeric and its value must be identical for all outlets within the same variance unit. If the variance unit is the same as the corresponding sampling PSU, this field must be the same as the value of Column E. For the example above, the 10 records from Variance Unit 101 would have this field filled by 34, those from Variance Unit 102 by 27, and those from Variance Unit 103 by 20. If two or more PSUs have been combined to form a variance unit, then the total number of outlets for the variance unit must equal the total number of outlets in the combined PSUs that were used to form the variance unit.

Column J - Sampling Weight

The overall base weight of the outlet (i.e., the inverse of the overall probability of outlet selection). The sampling weight is the unadjusted weight that can be calculated at the time of sampling. Adjustment for eligible noncomplete outlets will be done by the system within variance units.

Column K - Response Disposition Code

A code to indicate the final response disposition of the outlet. This field must be alphanumeric exactly as specified in Table 3. Either upper or lower case for the character part of the codes may be used. If, for any record, this field was coded as "other" (i.e., N9 or I10), then the user should give the "other" reasons and counts in the space provided in SSES output Table 3 (Synar Survey Sample Tally Summary).

Column L - Violation Flag

An indicator of whether or not the outlet was found to be in violation. This field must have a value of 0 or 1. Outlets found to be in violation must have this flag set to "1". Outlets found to be in compliance must have this field set to "0". Ineligible or eligible noncomplete outlets cannot have this field set to "1". The violation codes for inspection results are shown in Table 4, which is reproduced below.

Violation codes

Inspection result	Violation code
Outlet found to be in violation	1
Outlet found to be in compliance	0 or blank
Eligible noncomplete outlet	blank
Ineligible outlet	blank

Column M - Total Outlets in Sampling Stratum

The total number of outlets in the sampling frame that are in the sampling stratum. This field must be numeric if present and the same within sampling stratum, however it may be blank if not available (for example, if area sampling was used). Information provided in this column is used to fill out the outlet frame size column of SSES output Table 2 (see Exhibit 3).

Column N – Outlet Type

An indicator of the type of the outlet. There are two types of outlets: over-the-counter coded as "OTC" and vending machine, coded as "VM". If the type is unknown, it must be coded as "UNK" since a missing outlet type is not allowed. When a record with an unknown outlet type is present in the data file, a special message will be printed at the bottom of SSES output Table 2 indicating that the overall counts may not be equal to the sum of OTC and VM counts.

Column O – Youth Inspector ID

A unique identifier for each youth inspector. This field can be numeric or character. Each inspector must have a unique ID, and for a given inspector ID, inspector characteristics such as gender and age must be consistent. The system does not utilize this field if the outlet record has a response disposition code of ineligible or eligible noncomplete.

Column P – Youth Inspector Gender

This field should be coded with the youth inspector's gender, either "M" for male, "F" for female, or blank if unknown. Other values are not allowed. If the gender is unknown but the age is within the allowable range, then the inspector will be categorized as Other and the user should provide explanation for the "Other" category in SSES output table 4.

Column Q – Youth Inspector Age

Youth inspector age should be numeric or blank if unknown. If a sample outlet is eligible and has a complete inspection (i.e., the record disposition code is "EC"), then the age must be greater than or equal to 14 but less than or equal to 18. If the age is outside of this range or unknown, the outlet should be declared as eligible noncomplete because the inspection is invalid, and then assigned the disposition code "N9", for "Other" noncompletion. The user should provide this specific reason in SSES output Table 4. For a given Youth Inspector ID, gender and age must be consistent across all records; otherwise, an error message will appear and the system will not run unless the error is corrected by updating the data.

Column R – Vending Machine Outlet Frame Size in Stratum

This field provides the total number of vending machine outlets in the sampling frame that are in the stratum. This field must be numeric or blank, and its value must be identical for all outlets

within the same sampling stratum. If the number of vending machines is unknown or no vending machine is selected, leave it blank. If it is left blank, it will be treated as zero.

7.   Interpreting the Output

The SSES system produces four output tables, each of which occupies one Excel worksheet of a single workbook. Each table is discussed separately.

7.1   SSES Table 1 (Synar Survey Estimates and Sample Sizes)

The first worksheet, titled "SSES Table 1 (Synar Survey Estimates and Sample Sizes)" is shown in Exhibit 2. This table presents the main survey estimates and sample sizes.

SSES output Table 1 shows the following statistics:

a. *Unweighted* and *weighted retailer violation rates*;
b. *Standard error* of the weighted RVR;
c. *One-sided* and *two-sided 95 percent confidence intervals* of the weighted RVR;
d. Statement about whether or not the SAMHSA precision requirement has been met - the criterion used in this statement is the 3 percent margin of error for the one-sided 95 percent confidence interval;
e. *Design effect*;
f. *Weighted* and *unweighted accuracy rates* (an estimate of the proportion of outlets on the frame that are actually eligible for the Synar survey);
g. *Completion rate* (the proportion of eligible sampled outlets that were actually inspected with a clear result either in violation or compliance).

SSES output Table 1 also reports the sample sizes and the overall sampling rate for the current year. The table reproduces the effective and target sample sizes provided in the pop-up window at the beginning of your SSES session. The remaining sample sizes are generated by the system based on the sample data. Each of the sample sizes is described below:

a. *Effective sample size*: the sample size needed to meet the SAMHSA precision requirement under simple random sampling;
b. *Target (minimum) sample size*: the sample size to be achieved, which can be obtained as the product of the effective sample size and the design effect;

c. *Original sample size*: an inflated sample size of the target to account for sample attrition due to ineligibility and non-completion;

d. *Eligible sample size*: the number of outlets found to be eligible in the sample;

e. *Final sample size*: the number of eligible outlets in the sample for which inspection was completed.

Exhibit 2.   SSES Output Table 1 - example

7.2   SSES Table 2 (Synar Survey Results by Stratum and by OTC/VM)

The second worksheet titled "SSES Table 2 (Synar Survey Results by Stratum and by OTC/VM)" is presented in Exhibit 3. This table contains similar information to that provided in the new ASR Form 1, the optional Form 2, and Form 3 if a cluster design is used. The following information is provided for the overall universe, by outlet type (OTC and VM), and by sampling stratum:

a.   Outlet frame size;

b.   Estimated (eligible) outlet population size;

c.   Number of PSU clusters created (only for a cluster design);

d.   Number of PSU clusters in sample (only for a cluster design);

e.   Outlet sample size;

f.   Number of eligible outlets in sample;

g.   Number of sample outlets inspected;

h.   Number of sample outlets in violation;

i.   Retailer violation rate;

j.   Standard error (only for the aggregate RVR).

If the VM outlet frame size for a sampling stratum (Column R in the input data file) was left blank, the system assumes zero VM in the stratum. If some outlets have unknown outlet type, the sum of estimated population sizes for OTC and VM is not equal to the "All Outlets" count for the affected sampling stratum. When this happens, a message appears at bottom of the table warning you about potential inconsistencies in the table.

For a stratified cluster design with a list sampling frame, the frame size in the third column of the table is calculated using the information (total outlets in sampling stratum) provided in Column M in the data file. The estimated outlet population size column provides estimated total numbers of eligible outlets for the Synar survey by sampling stratum. The estimates are produced based on the sampling weights and the sample data you provided without using the frame size in Column M in the data file. If an area frame is used, this information is not available and Column M will be blank, and therefore, zeros will appear in the frame size column of the table. If a list-assisted area frame is used, this information is available but may be very unreliable and could be very different from the estimated outlet population size.

Exhibit 3.   SSES Output Table 2 - example

Samp. Stratum	Var. Stratum	Outlet Frame Size	Estimated Outlet Population Size	Number of PSU Clusters Created	Number of PSU Clusters in Sample	Outlet Sample Size	Number of Eligible Outlets in Sample	Number of Sample Outlets Inspected	Number of Sample Outlets in Violation	Retailer Violation Rate(%)	Standard Error(%)
						All Outlets					
1	1	814	632	10	3	215	205	199	15	6.3%	
2	2	670	664	20	7	241	235	227	23	10.1%	
3	3	550	529	103	21	159	154	149	16	14.1%	
Total		2,034	1,825	133	31	615	594	575	54	10.0%	1.8%
						Over the Counter Outlets					
1	1	788	628	10	3	212	204	198	14	5.6%	
2	2	635	656	20	7	238	232	224	21	9.5%	
3	3	530	518	103	21	154	149	144	15	14.0%	
Total		1,953	1,802	133	31	604	585	566	50	9.4%	1.8%
						Vending Machines					
1	1	26	4	10	3	3	1	1	1	100.0%	
2	2	35	8	20	7	3	3	3	2	66.7%	
3	3	20	11	103	21	5	5	5	1	20.0%	
Total		81	23	133	31	11	9	9	4	50.4%	16.2%

STATE: XX
FFY: 2005

7.3   SSES Table 3 (Synar Survey Sample Tally Summary)

The third worksheet titled "SSES Table 3 (Synar Survey Sample Tally Summary)" (see Exhibit 4) is a summary of final response disposition codes for the survey. The total number of eligible and inspection complete outlets is given, along with the number of eligible, noncomplete outlets, and ineligible outlets. Separate tallies are given by reason for noncompletion or ineligibility. This table contains the same information as that provided in ASR Form 4. If there are any outlets that fall into the "Other" categories (coded "N9" or I10"), a space will be automatically provided at the bottom of the table where each "Other" reason and the counts for each reason are to be provided.

Exhibit 4. Sample SSES Output Table 3 - example

	Disposition Code	Description	Count	Subtotal
		SSES Table 3 (Synar Survey Sample Tally Summary) STATE: XX FFY: 2005		
5	EC	Eligible and inspection complete outlet	575	
6	Total (Eligible Completes)			575
7	N1	In operation but closed at time of visit	5	
8	N2	Unsafe to access	6	
9	N3	Presence of police	3	
10	N4	Youth inspector knows salesperson	2	
11	N5	Moved to new location but not inspected	3	
12	N6	Drive thru only/youth inspector has no drivers license	0	
13	N7	Tobacco out of stock	0	
14	N8	Run out of time	0	
15	N9	Other noncompletion	0	
16	Total (Eligible Noncompletes)			19
17	I1	Out of Business	14	
18	I2	Does not sell tobacco products	1	
19	I3	Inaccessible by youth	3	
20	I4	Private club or private residence	1	
21	I5	Temporary closure	2	
22	I6	Unlocatable	0	
23	I7	Wholesale only/Carton sale only	0	
24	I8	Vending machine broken	0	
25	I9	Duplicate	0	
26	I10	Other ineligibility	0	
27	Total (Ineligibles)			21
28	Grand Total			615

7.4   SSES Table 4 (Synar Survey Inspection Results by Youth Inspector Characteristics)

The fourth worksheet titled "SSES Table 4 (Synar Survey Inspection Results by Youth Inspector Characteristics)" is presented in Exhibit 5. In the first part of the "Table 4", the number of inspectors, attempted buys and successful buys are given by gender and age of inspectors. The second part of the "Table 4" gives the buy rates calculated by gender and age of inspector. Note that if gender is

not reported for some inspectors, their counts and buy rates will be given in the "Other" row. This table reports the information needed in Form 5 of the ASR.

Exhibit 5.   Sample Summary Table 4 - example

8.   What to Include in the Annual Synar Report

SSES is an optional tool designed to assist States in analyzing and reporting their annual Synar survey results. By simply attaching the four output tables to the ASR, most of reporting requirements for the Synar survey results and analyses will be met. This will greatly reduce the reporting burden and expedite the review process. However, if data are not complete or discrepancies are noted,

additional information may be required to complete the report. The equivalency of SSES Tables and the relevant reporting requirements for the ASR is shown table 5:

Table 5.   Equivalency of SSES Tables and the ASR

SSES Table	ASR
1	Questions 7b – 7g
2	Form 1, Form 2, Form 3
3	Form 4
4	Form 5

All States using the SSES must report the remaining sections of the ASR.

9.   Handling of Special Cases

Problems can be caused by poor field operation or unusual circumstances such as natural disasters during the Synar field work and sometimes by the sample design itself (e.g., a single PSU is selected per stratum). This section addresses such special situations. It also deals with some standard techniques, such as post-stratification, available for non-anomalous situations to enhance the sampling efficiency or the usefulness of the Synar survey data.

9.1   Requirement for Nonresponse Adjustment

In producing the RVR estimate, the SSES always performs a nonresponse adjustment to account for eligible non-inspected outlets. This adjustment is done within variance strata for the Stratified SRS option and within variance units for the Stratified CLUSTER option. Therefore, there must be at least one responding (eligible complete) outlet in every variance stratum or unit where nonresponse adjustment is performed. If any stratum or unit has zero responding outlets, SSES will give an error message informing the user to combine such variance strata or units with others. This can be done by redefining the variance strata or variance units in the data file.

9.2   Combining Sampling Strata into a Variance Stratum

Normally, at least two sampling units per stratum are needed to be able to estimate the variance of the estimated RVR; otherwise a problem can arise. Anomalous situations and their remedies for variance estimation are explained separately for different sample design options.

For the Stratified SRS, suppose that there is only one eligible responding outlet in a sampling stratum. If this is the only outlet in the sampling frame for that sampling stratum (i.e., the outlet was selected with certainty), then the variance contribution from the stratum is well defined and equal to zero and so, it is not a problem. However, if there are other outlets in the frame for the sampling stratum, then the variance estimator from the stratum with only one outlet is undefined. To estimate the variance in this situation, it is necessary to combine the sampling stratum with other sampling stratum to form a variance stratum that contains more than one eligible responding or ineligible outlet. Note that ineligible outlets are counted in the sample for variance estimation purposes but not eligible noncomplete outlets. When combining of sampling strata occurs, the Columns D (Variance Stratum) and E (Total Outlets in Variance Stratum) in the input data file must be properly modified for the affected outlet records. To gain the most benefit from stratification, combine strata that are expected to be similar, particularly with respect to stratum sampling rates and their retailer violation rate.

For the Stratified CLUSTER, the variance of the estimated RVR consists of two components: between-PSU variance and within-PSU variance. In order to estimate the variance, both components should be well defined. Consider a situation in which there is only one sampled PSU in a sampling stratum. If this is the only PSU in the sampling frame for that sampling stratum (i.e., the PSU was selected with certainty), then the between-PSU variance component for the stratum is well defined and equal to zero, and, thus, it is not a problem as long as the within-PSU variance component is well defined. However, if the PSU has a single outlet, it can cause problems in calculating the within-PSU variance component (discussed in Section 9.3). If the single sample PSU was selected among other PSUs in the sampling stratum, then the between-PSU variance component cannot be estimated. A remedy for the problem is to combine the sampling stratum with other sampling stratum to form a variance stratum that contains more than one sample PSU. When combining of sampling strata occurs, the columns F (Variance Stratum) and G (Total Variance Units in Variance Stratum) in the input data file must be properly modified for the affected outlet records. To gain the most benefit from stratification, combine strata that are expected to be similar, particularly with respect to stratum sampling rates and their retailer violation rate.

9.3   **Combining Sampling PSUs into a Variance Unit**

This section is relevant only when the Stratified CLUSTER option is used.

Suppose that there is only one eligible responding outlet in a sampling PSU. If this is the only outlet in the sampling frame from that sampling PSU, then the variance contribution from the PSU (within-PSU variance) is well defined and equal to zero. However, if there are other outlets in the frame, then the variance from this PSU is undefined for the With-FPC option. To estimate the variance in this situation, it is necessary to combine the sampling PSU with other sampling PSUs in the same sampling stratum, as much as possible, to form a variance unit that contains more than one eligible responding or ineligible outlet. However, when the Without-FPC option is used, single outlet PSUs do not cause a problem because the variance estimate is defined by the between-PSU variance only.

9.4   **Certainty PSUs**

Certainty PSUs are primary clusters selected with probability equal to 1. This is common phenomenon with a cluster design that selects PSUs with unequal probabilities, especially when the PSUs are very different in size within strata. The correct way to handle certainty PSUs in variance estimation depends on the number of stages of clustering in the design.

Case 1 – one- or two-stage cluster design: create a variance stratum that consists of all certainty PSUs, and then define the PSUs as variance units for the variance stratum. In this case, the total variance units in the variance stratum (Column G of the input file) must be equal to the number of certainty PSUs in the sample (they can be pooled across the sampling strata). The effect of this makes the between-PSU variance zero and the variance from the certainty PSUs is solely from the within-PSU variance.

Case 2 – three- or higher-stage cluster design: each of certainty PSUs forms its own variance stratum and the secondary clusters within certainty PSUs become variance units. The Without-FPC option must be used. The resulting variance estimate will be conservative but the conservatism will usually be mild since such a complex design is usually used when sampling from a large population with a small sampling rate.

9.5   **Vending Machines**

If vending machines were selected separately from over-the-counter outlets, then all vending machines should be placed in one or more separate variance stratum or strata of their own. Whether

vending machines were selected separately or together with over-the-counter outlets, the new system produces the violation rates by outlet type in SSES output Table 2, as long as the outlets are correctly categorized as OTC or VM in the input data file.

9.6   Post-stratification

Post-stratification is a technique often used in survey sampling to create estimation strata, which are different from the sampling strata, after sampling was done. It is often beneficial in reducing the sampling variance. However, it requires population information (frame size) for post-strata. For example, suppose that a state-wide SRS was selected and the sampling frame can provide frame counts for the state's geographic regions. The state can use the regions as post-strata, which should be turned into SSES variance strata to accomplish the post-stratification. If the state wants to produce RVR estimates by geographic region, this can be accomplished by defining the post-strata as both sampling strata and variance strata when creating the input data file. Then the system produces RVR estimates by region in SSES output Table 2. If this is done, it is important to explain the redefinition of the sampling strata in the ASR; otherwise, it would not be noticed because the post-strata would appear just like design strata.

Another use of poststratification is to reflect implicit stratification in variance estimation when systematic sampling is used. This can be done by making implicit stratification explicit through poststratification. For example, suppose that in a sampling stratum with 825 outlets, 100 outlets were selected systematically after sorting the stratum sampling frame in a certain order. Preserving this order, the stratum is divided into four subgroups of more or less equal size. These subgroups can be used as post-strata, which will be referred as PS1, PS2, PS3, and PS4. Let PS1 have 207 outlets, and the other post-strata all have 206 each, in the frame. Suppose further that the sample of 100 outlets is distributed over these post-strata by 23, 27, 30, and 20. Then, define the four post-strata as variance strata. In the data file, the 100 sample outlets should all have the same sampling stratum ID and the same sampling stratum population size of 825, but each would have one of four different variance stratum ID depending on which post-stratum the outlet belongs to – the 23 outlets in PS1 will have the variance stratum ID "PS1" and the variance stratum population size of 207, the 27 outlets in PS2 will have variance stratum ID "PS2" and the variance stratum population size of 206, and so forth.

In the case of cluster design, the same principle can be used to define post-strata (i.e., variance strata) to reflect implicit stratification of PSU clusters. Note that, however, this technique does not reflect the implicit stratification to the fullest extent because it is impossible to turn the implicit stratification into explicit stratification perfectly. The resulting variance estimate will still be somewhat

conservative. If post-stratification is not used, it will be even more conservative than a variance estimate obtained from some post-stratification.

9.7   Cluster Sample Design with More Than Two Stages

Most Synar surveys that employ a cluster design use a one- or two-stage design. SSES is particularly suitable for such designs. SSES can also be used for cluster designs with more than 2 stages by selecting the Without-FPC option. This produces a mildly conservative variance estimate. If the overall sampling rate is small (which is usually the case in practice when a complex design is used), this is acceptable. If there are certainty PSUs selected, see Section 9.4 for proper handling of the certainty PSUs.

To be able to use the With-FPC option for such a design requires the calculation of variance components of higher than second stage sampling. This would be unduly complicated, and that is why most commercial survey sampling software packages assume a cluster design with no more than two stages.

10.   SSES Variance Estimation Approach

There are two broad approaches to variance estimation for sample surveys: Taylor linearization method and replication (or resampling) method. The SSES uses the former approach. The system uses the domain estimation technique to estimate the variance (or standard error) of the weighted retailer violation rate. This technique takes into account the fact that list frames used in the survey always contain some ineligible outlets. In this case, the population of eligible outlets is a domain of the larger population that is represented by the list sampling frame. Because of this, the sample drawn from the frame contains some ineligible outlets of unspecified number before sampling, which makes the sample size of eligible outlets random – the randomness here means that the sample size of eligible outlets is unpredictable before sampling, even though the overall sample size is fixed. This randomness of the sample size is reflected in the variance estimation by the domain estimation technique. However, this technique affects only the variance estimation, not the estimation of the weighted retailer violation rate.

11.   Technical Support

For technical support, contact:

Alejandro A. Arias, Ed.D.
Government Project Officer,
SAMHSA/CSAP

301-443-4825

AArias@samhsa.gov.

Appendix A: Error and Warning Messages and Troubleshooting

The system is designed to provide various error and warning messages to help the user correct the problem. However, the system does not check every possible problem or data inconsistency. It is, therefore, ultimately the user's responsibility to prepare the data file correctly and review the internal and external consistencies of the data. Much of this Guide has been devoted to assisting the user to accomplish this.

In this section, the "Stratified SRS" and "Stratified CLUSTER" options will be referred simply as "SRS" and "CLUSTER" respectively. Three tables in the section summarize the error and warning messages that may occur for different menu options and suggests what course of action to take: Table 6 for the SRS option, Table 7 for the CLUSTER option, and Table 8 for the Sample Size Calculator.

Many of the error messages shown in Tables 6, 7, and 8 refer to specific Excel worksheet row numbers and column letters, such as Row 3, Column D. The tables present the message using the following generic format:

a. Column [c] represents a generic column, where c stands for Excel worksheet column designation (e.g., A, B, C, etc.);

b. Row [r] represents a generic row, where r stands for Excel worksheet row designation (e.g., 1, 2, 3, etc.);

c. [*column name*] is the column heading that the user defined in the first row of the Excel data file;

d. (*character*) refers to the non-numeric value contained in the cell of the Excel data file. In SSES, the actual non-numeric value contained in the cell will be displayed in the error message;

e. (*number*) refers to the numeric value contained in the cell. In SSES, the actual number contained in the cell will be displayed in the error message;

f. (*value*) can either be a numeric or an alphanumeric value contained in the cell. In SSES, the actual value contained in the cell will be displayed in the error message.

Table 6.　　Error messages summary for the Stratified SRS option

Error or warning message	Action
Invalid or empty column name in Row 1, Column [*c*].	Make sure the user defined column heading in Column [*c*] is valid and not missing.
You have selected the 'Stratified SRS' option but the data set does not match that option. Please select an option compatible with the data.	This message is given when the user tries to run the SRS option with a data file laid out for the CLUSTER option. The system recognizes the data file as for the CLUSTER option if the column title in Column M and Row 1 is given. Check the data file or select the correct option.
The value for [*column name*] in Row [*r*] is missing.	Missing values are not allowed in some columns in the data file. This message indicates that the outlet record given in Row [*r*] has a missing value for the designated column labeled as [*column name*] by the user. The columns that must not be missing are A, B, C, D, E, F, and H.
The value (*character*) for [*column name*] in Row [*r*] is not a number.	The value given for the cell in the column referenced and Row [*r*] is not numeric when a number is expected. Enter an appropriate number in the cell.
Values in the [*column name*] column are not unique. The following values were found: (*value*) (in Row: [*r1, r2*]) If correct press 'OK' to continue, otherwise press 'Cancel' to abort.	The same outlet ID (*value*) in Column A labeled as [*column name*] appears more than once. Make sure that outlet ID's are unique for each outlet in the sample. This is to prevent duplicate records in the data file.
Invalid value (*value*) for [*column name*] in Row [*r*]. Enter an admissible value.	If column referenced is Column F, make sure that response disposition code is a valid code in Row [*r*].
Invalid value (*value*) for [*column name*] in Row [*r*]. Expecting 0, 1 or [blank].	Make sure that the value in Column G labeled as [*column name*] and Row [*r*] is 0, 1, or [blank].
Invalid value (*value*) for [*column name*] in Row [*r*]. Expecting OTC, VM or UNK.	Make sure that the value in Column H labeled as [*column name*] and Row [*r*] is "OTC", "VM" or "UNK".
Values of [*column name1*] must be the same within [*column name2*]. Inconsistent value (*number*) was found in Row [*r1*] and Row [*r2*].	If column referenced is Column C (column heading given by [*column name1*]), make sure that the frame size of the sampling stratum (column heading given by [*column name2*]) is identical for all outlets within the same sampling stratum.
Values of [*column name1*] must be the same within [*column name2*]. Inconsistent value (*number*) was found in Row [*r1*] and Row [*r2*].	If column referenced is Column E (column heading given by [*column name1*]), make sure that the frame size of the variance stratum (column heading given by [*column name2*]) is identical for all outlets within the same variance stratum.
Values of [*column name1*] must be the same within [*column name2*]. Inconsistent value (*number*) was found in Row [*r1*] and Row [*r2*].	If column referenced is Column L (column heading given by [*column name1*]), make sure that the vending machine frame size of the sampling stratum (column heading given by [*column name2*]) is identical for all outlets within the same sampling stratum.

Table 6. Error messages summary for the Stratified SRS option (continued)

Error or warning message	Action
When the definition of [Variance strata] and [Sampling strata] are equal, the values in [*column name1*] and [*column name2*] should also be equal. Values (*value1*, *value2*) found in Row [*r*] are not equal. Please check and rerun.	When Column D for the variance stratum ID and Column B for the sampling stratum ID are the same, make sure the values in Column E labeled as [*column name1*] and Column C labeled as [*column name2*] are equal.
The population size for variance stratum (*value*) is smaller than the sample size, correct and rerun.	The sample size (the number of outlet records) in a variance stratum identified by (*value*) should be smaller than or equal to the values in Column E for the variance stratum.
The value in [*column name1*] in Row [*r*] is incompatible with the value in [*column name2*].	Make sure that the value in Column G labeled as [*column name1*] and Row [*r*] is blank or "0" if the value in Column F labeled as [*column name2*] is not "EC".
Only one or no respondent in variance stratum (*value*). Please combine the variance stratum with other and rerun.	Make sure that there are more than 1 responding outlets in the variance stratum identified by (*value*). If not, see Section 9.2.
No respondent in variance stratum (*value*). Please combine the variance stratum with other and rerun.	Make sure that there are some responding outlets in the variance stratum identified by (*value*). If not, see Section 9.2.
Inspector ID is missing in Row [*r*]. Please enter an ID and rerun.	Make sure that Column I for the youth inspector ID in Row [*r*] is not missing. This error will be triggered if it is missing and the response disposition code in Column F for the record is "EC".
Inspector Age value (*value*) in Row [*r*] is outside the permissible range (14-18). If the age is correct, then the record must have [*column name*] other than "EC". If the age is not correct, correct it and rerun.	Make sure that Column K for youth inspector age in Row [*r*] is in the range of 14-18 for the eligible complete outlet. If the age is correct, then the record must be coded as "N9" in Column F labeled as [*column name*].
Inconsistent Gender values for inspector ID = (*ID value*) Row [r1] ID: (*ID value*), AGE: (*age value*), GENDER: (*gender value1*) Row [r2] ID: (*ID value*), AGE: (*age value*), GENDER: (*gender value2*)	Make sure that Column J for the youth inspector gender is consistent in Rows [*r1*] and [*r2*] for the inspector with ID given by (*ID value*).
Inconsistent Age values for inspector ID = (*ID value*) Row [*r1*] ID: (*ID value*), AGE: (*age value1*), GENDER: (*gender*) Row [r2] ID: (*ID value*), AGE: (*age value2*), GENDER: (*gender*)	Make sure that Column K for the youth inspector age is consistent in Rows [*r1*] and [*r2*] for the inspector with ID given by (*ID value*).

Table 7. Error messages summary for the Stratified CLUSTER option

Error or warning message	Action
Invalid or empty column name in Row 1, Column [*c*].	Make sure the column heading in Column [*c*] is valid and not missing. Make sure the user defined column heading in Column [*c*] is valid and not missing.
Invalid or empty column name in Row 1, column M, or a Stratified CLUSTER option is selected for Stratified SRS data. Correct and rerun.	This message is given when the user tries to run the CLUSTER option with data file for the SRS option. The system recognizes the data file as for the SRS option if the column title in Column M and Row 1 is empty. Check the data file or select the correct option.
The value for [*column name*] in Row [*r*] is missing.	Missing values are not allowed in some columns in the data file. This message indicates that the outlet record given in Row [*r*] has a missing value for the designated column labeled as [*column name*] by the user. The columns that must not be missing are A, B, C, D, E, F, G, H, I, J, K, and N.
The value (*character*) for [*column name*] in Row [*r*] is not a number.	The value given in the column referenced and Row [r] is not numeric when a number is expected. Enter an appropriate number in the cell.
Values in the [*column name*] column are not unique. The following values were found: (*value*) (in Row: [*r1, r2*]). If correct press 'OK' to continue, otherwise press 'Cancel' to abort.	The same outlet ID (*value*) in column A labeled as [*column name*] appears more than once. Make sure that outlet ID's are unique for each outlet in the sample. This is to prevent duplicate records in the data file.
Invalid value (*value*) for [*column name*] in Row [*r*]. Enter an admissible value.	This error occurs when a nonadmissible value is used in Column K labeled as [*column name*] for the response disposition code. Make sure that response disposition code is a valid code in Row [*r*].
Invalid value (*value*) for [*column name*] in Row [*r*]. Expecting 0, 1 or [blank].	Make sure that Column L labeled as [*column name*] has only 0, 1, or [blank] in Row [*r*].
Invalid value (*value*) for [*column name*] in Row [*r*]. Expecting OTC, VM or UNK.	Make sure that the value in Column N labeled as [*column name*] and Row [*r*] is "OTC", "VM" or "UNK".
Invalid value (*value*) for [*column name*] in Row [*r*]	If the column referenced is Column J labeled as [*column name*], make sure that the sampling weight is numeric and positive.
Values of [*column name1*] must be the same within [Sampling stratum]. Inconsistent value (*number*) was found in Row [*r1*] and Row [*r2*].	The values given in column labeled as [*column name1*] must be consistent within the sampling stratum but they are not in Rows [*r1*] and [*r2*]. Correct the wrong values and rerun.
Values of [*column name1*] must be the same within [Variance stratum]. Inconsistent value (*number*) was found in Row [*r1*] and Row [*r2*].	The values given in column labeled as [*column name1*] must be consistent within the variance stratum but they are not in Rows [*r1*] and [*r2*]. Correct the wrong values and rerun.

Table 7. Error messages summary for the Stratified CLUSTER option (continued)

Error or warning message	Action
When the definition of [Variance stratum] and [Sampling stratum] are equal, the values in [*column name1*] and [*column name2*] should also be equal. Values found in Row [*r*] are not equal. (*value1*)<>(*value2*) Please correct and rerun.	When the variance strata given in Column F and the sampling strata given in Column B are the same, the values in column labeled as [*column name1*] and column labeled as [*column name2*] should also be equal but they are not in Row [*r*]. Correct the error and rerun.
When the definition of [Variance Stratum] and [Variance Unit] are the same as [Sampling Stratum] and [Sampling PSU], the values in [*column name1*] and [*column name2*] should be equal. Values found in Row [*r*] are not equal. (*value1*) <> (*value2*) Please check and rerun.	When the variance strata given in Column F and the sampling strata given in Column B are the same, and the variance units in Column H labeled as [*column name1*] are the same as the sampling PSUs in Column D labeled as [*column name2*], the values in column labeled as [*column name1*] and column labeled as [*column name2*] should be equal but they are not in Row [*r*]. Correct the error and rerun.
Total number of frame variance units (*value1*) given in variance stratum (*ID*) is less than the number of variance units in the sample (*value2*). Please correct and rerun.	The total number (*value2*) of sample PSUs in the variance stratum identified by (*ID*) should be smaller than or equal to the value (*value1*) in Column G. Correct the value (*value1*) and rerun.
Total number of frame outlets (*value1*) given in variance unit (*ID1*) of variance stratum (*ID2*) is less than the number of sample outlets in the variance unit (*value2*). Please correct and rerun.	The total number (*value2*) of sample outlets in the variance unit identified by [*ID1*] of the valiance stratum identified by [*ID2*] should be smaller than or equal to the value (value1) in Column I. Correct the value (*value1*) and rerun.
The value in [column name1] in Row [*r*] is incompatible with the value in [*column name2*].	Make sure that the value in Column L labeled as [*column name1*] and Row [*r*] is blank or "0" if the value in Column K labeled as [*column name2*] is not "EC".
Only one or no respondent in variance unit (*ID1*) of variance stratum (*ID2*). Please combine the variance unit (*ID1*) with other and rerun.	Make sure that there are at least two eligible responding and/or ineligible outlets in the variance unit identified by (*ID1*) of the variance stratum identified by (*ID2*) for the With-FPC option (or one such outlet if it is the only outlet in the frame in that variance unit). If not, see section 9.3.
No variance unit with respondent in variance stratum (*ID*).	Make sure that there is at least one variance unit containing an eligible responding and/or ineligible outlet in the variance stratum identified by (*ID*). If not, see section 9.2.
Only one variance unit in variance stratum (*ID*). Please combine the variance stratum and rerun.	Make sure that there are at least two variance units in each variance stratum. If not, see section 9.2.
No respondent in variance unit (*ID1*) of variance stratum (*ID2*). Please combine the variance unit (*ID1*) with other and rerun.	Make sure that there is at least one eligible responding outlet in the variance unit identified by (*ID1*) of the variance stratum identified by (*ID2*). If not, see section 9.3.

Table 7. Error messages summary for the Stratified CLUSTER option (continued)

Error or warning message	Action
Inspector ID is missing in Row [*r*]. Please enter an ID and rerun.	Make sure that Column O for the youth inspector ID in Row [*r*] is not missing. This error will be triggered if it is missing and the response disposition code in Column K for the record is "EC".
Inspector Age value (*value*) in Row [*r*] is outside the permissible range (14-18). If the age is correct, then the record must have [*column name*] other than "EC". If the age is not correct, correct it and rerun.	Make sure that Column Q for youth inspector age in Row [*r*] is in the range of 14-18 for the eligible complete outlet. If the age is correct, then the record must be coded as "N9" in Column K labeled as [*column name*].
Inconsistent Gender values for inspector ID = (*ID value*) Row [*r1*] ID: (*ID value*), AGE: (*age value*), GENDER: (*gender value1*) Row [*r2*] ID: (*ID value*), AGE: (*age value*), GENDER: (*gender value2*)	Make sure that Column P for the youth inspector gender is consistent in Rows [*r1*] and [*r2*] for the inspector with ID given by (*ID value*).
Inconsistent Age values for inspector ID = (*ID value*) Row [*r1*] ID: (*ID value*), AGE: (*age value1*), GENDER: (*gender*) Row [r2] ID: (*ID value*), AGE: (*age value2*), GENDER: (*gender*)	Make sure that Column Q for the youth inspector age is consistent in Rows [*r1*] and [*r2*] for the inspector with ID given by (*ID value*).

Table 8. Warning message summary for the sample size calculator

Warning message	Action
Please verify your entry for Frame Size (*value*)	The Frame Size should be a positive number. Otherwise this message appears.
Please verify your entry for Safety Margin (*value*)	The Safety Margin should be between 0 and 50. Otherwise this message appears.
Please verify your entry for RVR (*value*)	RVR should be between 0 and 100. Otherwise this message appears.
Please verify your entry for Accuracy Rate (*value*)	The Accuracy Rate should be between 50 and 100. Otherwise this message appears.
Please verify your entry for Completion Rate (*value*)	The Completion Rate should be between 50 and 100. Otherwise this message appears.
Outside Range. Design effect is outside the range of (.5, 5), continue anyway?	Make sure the design effect is correct. If you choose to continue, the system will use whatever value provided.

Appendix B: Frequently Asked Questions (FAQ)

B.1   **About SSES**

1. What is SSES?

    SSES stands for Synar Survey Estimation System. This is a software package developed in a language that can be run in Microsoft Excel in the PC environment. It is designed to assist States in analyzing and reporting their annual Synar survey results.

2. What is the benefit of using SSES?

    Because it is a tailor-made system solely for the Synar survey, it is less cumbersome and more convenient for the Synar survey analysts than general statistical analysis or database software. SSES combines data entry, analysis, and reporting together. It produces all data items that should be produced and reported from the Synar survey and, thus, it simplifies the reporting and reviewing process, and enhances standardization of data collection, processing, and reporting. It is freely distributed and technical supports are readily available to the users.

3. Who developed the SSES?

    SSES is developed by the Center for Substance Abuse Prevention (CSAP) to assist States and Jurisdictions in conducting the Synar survey, and in analyzing and reporting the survey results.

4. How can I obtain technical assistance for SSES?

    For technical assistance for the system, contact Dr. Alejandro Arias, Project Officer, CSAP, by telephone at 301-443-4825 or by email at AArias@samhsa.gov.

5. Is it mandatory for the states and jurisdictions to use SSES for the Annual Synar Report (ASR)?

    No, it is optional but highly recommended for the benefit of everyone involved in the program.

6. What is the difference between SSES version 2.1 and SSES version 1.1?

    In the new versions, there are three main differences: (1) the table for Synar survey inspection results by youth inspector characteristics is generated; (2) the retailer violation rate (RVR) is estimated by outlet types (over-the-counter/vending machine) as well as overall; (3) the sample size calculator is a separate module that can run independently from the analysis options. The focus of the new version 2.1 is to produce

all required data items or information that are required in the ASR, summarizing the results from the Synar survey. The general format and data checking was also vastly improved.

B.2   **System Requirements**

1. What version of Microsoft Excel is needed to use SSES?

    SSES is designed to operate in Microsoft Excel 97 or higher version in the PC environment.

2. Can I use software other than Microsoft Excel to prepare the input dataset such as ASCII file?

    Yes, as long as you can import or convert the file into a Microsoft Excel file in the exact format laid out in this manual. However, it is recommended to use Excel for data compilation.

3. Is there any difference in the output results when different versions of Excel are used?

    Virtually none. However, the appearance of the output tables may be slightly different.

4. Can SSES system affect my other Excel files?

    No, but it is recommended to save and close other Excel files before running SSES.

B.3   **Data Entry**

1. Can I use column names other than the ones given in the template or the example datasets?

    Yes, but it is recommended to use meaningful names.

2. Can I use names of geographic areas such as county or ZIP code as stratum or PSU ID?

    Yes, as long as they are unique for identification purposes.

3. Can I use sample outlet ID's used by the field staff as outlet ID in the SSES data file?

    Yes, as long as they are unique.

4. What if the youth inspector ID is missing on the inspection form?

   If the youth inspector ID is missing for a respondent outlet (i.e., eligible and inspection complete), SSES will issue an error message. You have to provide the youth inspector ID for every respondent outlet. If unknown, the outlet must be classified as eligible noncomplete using the response disposition code "N9" – other eligible noncomplete.

5. If I use SSES, do I have to attach a copy of the data file along with the output tables?

   It is not required but recommended.

6. Can I enter additional information onto the SSES Excel data file?

   Yes, but the usable columns depend on the analysis option you intend to use. For the Stratified SRS option, you can enter additional information in any empty columns starting from Column N but not on Column M, which must be left blank for the SRS option. For the Stratified CLUSTER option, you can enter additional information in any empty columns starting from Column S.

7. Can I insert blank rows or columns in the Excel data file?

   No. A blank row marks the end of the data file in SSES and, thus, if a blank row is inserted in the middle of the data file, the system will read up to that point ignoring the rest. Therefore, it is important to make sure that this does not happen. The system shows the number of records read before running the system so that the user can check whether the size of the data file read is correct or not. No blank column should be inserted among the designated columns, since this will generate an error condition, which will stop running the system.

B.4   Running SSES

1. Can I run SSES if I do not know the effective sample size and the target sample size?

   The effective sample size and the target sample size should be provided in the first pop-up window, which appears when you select an analysis option. This information is not used in analysis and so does not affect the analysis results. They will simply be reproduced in SSES output table 1. The system accepts any positive number but it is important to provide the correct effective sample size and the target sample size.

2. What criteria should I use when choosing between "With-FPC" and "Without-FPC"?

   Since most State Synar survey use sampling without replacement, the With-FPC option may be used in most cases. The With-FPC should be used for the stratified simple or systematic random sampling designs (see Section 4.2.1 of the user manual for more discussion). For a stratified cluster design with unequal probability selection of PSUs, the With-FPC option could lead to a liberal variance estimate (i.e., smaller than it should

be), and therefore, the choice should be made with caution. If the design has more than two stages, then the Without-FPC option should be used (for details, see Sections 4.2.2 and 9.7 of the user manual).

3. If I don't know the number of vending machines in the population, what should I do?

   If the frame does not contain the vending machine information, the column in the data file for the vending machine frame size in sampling stratum should be left blank (column L for the Stratified SRS option and column R for the Stratified Clustered option. This affects only the frame size calculation in the SSES output Table 2.

4. Can I cancel a run after it started?

   No. If you find some error in the data file, you simply discard the old output file and rerun the system with the corrected data file.

5. Can I run the SSES system multiple times in the same session?

   Yes. In each run SSES will create a new Excel workbook which contains the analysis results. To keep the results after the session, the workbook must be saved.

6. What shall I do if the standard error does not meet the SAMHSA precision requirement?

   Check your data file and the analysis option you chose. If everything is correct, then you should report exactly as SSES produced. The problem can be rectified only in the next cycle of the Synar survey by increasing the sample size and/or making the sample design more efficient.

B.5   About the Sample Size Calculator

1. What is the effective sample size?

   The effective sample size is the sample size needed to meet the SAMHSA precision requirement for the simple random sample (SRS) design. In most cases, the actual sample design is more complex than the SRS and it is necessary to inflate the effective sample size to reflect the complexity. The reason why the effective sample size is first calculated to determine the sample size for a complex sample design is because it is easier to translate the SAMHSA precision requirement into the sample size under SRS.

2. What is the target sample size and how is it related to the effective sample size?

   The target sample size is the sample size that is obtained by multiplying the effective sample size by the design effect, to reflect the complexity of the actual design other than SRS. It is the target to achieve at the end after the Synar survey field operation is finished.

3. What is the planned original sample size?

Nearly all surveys suffer some sample loss during the field operation due to various reasons such as ineligibility of some sample units and noncompletion of some eligible sample units. To counter these losses, the target sample size is inflated by using the accuracy and completion rates and further adjusted by incorporating a safety margin. The sample size determined through this process is called the planned original sample size. This is the sample size to be used to select a Synar sample.

B.6   Sample Design

1. What kind of sampling designs can SSES handle?

   Stratified SRS and Stratified CLUSTER designs with or without replacement sampling. These encompass all Synar sample designs currently used.

2. Can SSES incorporate finite population correction into the estimation procedure?

   Yes. However, it is only an approximate correction for stratified cluster designs with unequal probability selection of PSUs.

3. Is Sampling Stratum always the same as Variance Stratum?

   Generally, this is the case. However collapsing sampling strata may be needed if there is no or only 1 respondent sampling unit in a sampling stratum. This is accomplished by redefining the variance strata. For detail, see Section 9.2 in the user manual.

4. If collapsing over strata is necessary, will the system collapse strata automatically?

   No. You have to do it. See Section 9.2 in the user manual.

5. Is the Sampling PSU always the same as Variance Unit for the cluster design?

   Generally, that is the case. However, collapsing sampling PSU may be needed if there is no or only 1 respondent in a sampling PSU. For detail, see Section 9.3 in the user manual.

6. If collapsing over PSUs is necessary, will the system collapse PSUs automatically?

   No. You have to do it. See Section 9.3 in the user manual.

7. How should certainty PSUs be treated?

Certainty PSUs need a special treatment in SSES. The general principle is to make them their own variance strata or put them together in a variance stratum. See Section 9.4 of the user manual.

8. Can I use SSES for a three or higher-stage sample design?

Yes, but you have to use the Without-FPC option. For detail, see Section 9.7 of the user manual.